



Trabajo de Fin de Máster: Tema: Estrategias para el tratamiento de datos faltantes ("missing data") en estudios con datos longitudinales

Anexo de Planificación

Román Octavio Calafati
Máster Universitario en Bioinformática y Bioestadística
Área Bioestadística

Directora: Nuria Pérez Álvarez
Profesor Responsable de la Asignatura: Alexandre Sánchez Pla

24 de mayo de 2017



Esta obra está sujeta a una licencia de Reconocimiento -
NoComercial - SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Índice

1. Introducción al anexo de planificación	1
1.1 Planificación temporal calendarizada del Trabajo. Tareas, hitos y recursos.	2
1.1.1 Planificación temporal calendarizada.....	2
1.1.2 Seguimiento del proyecto: utilización del software xPlan	3
1.1.3 Días libres de acuerdo al calendario de la Comunidad de Madrid:.....	4
1.1.4 Diagrama de Gantt.....	5
1.1.5 Tareas e hitos	6
1.1.6 Recursos	7
2. Incidencias, riesgos y plan de contingencia	8
2.1. Riesgo 1: mal funcionamiento del ordenador	8
2.1.1. Plan de contingencia y mitigación	8
2.2 Riesgo 2: pérdida de datos (virus, malware)	8
2.2.1 Plan de contingencia y mitigación	8
2.3 Riesgo 3: Solapamiento con entregas en la asignatura Análisis Multivariante..	9
2.3.1 Plan de contingencia y mitigación	9
2.4 Riesgo 4: Carga de tiempo extra en el trabajo	9
2.4.1 Plan de contingencia y mitigación	9
2.5 Riesgo 5: Viaje imprevisto	9
2.5.1 Plan de contingencia y mitigación	9
2.6 Riesgo 6: Alcance del proyecto	9
2.6.1 Plan de contingencia y mitigación	9
2.7 Riesgo 7: Problemas de salud personales inesperados	10
2.7.1 Plan de contingencia y mitigación	10

Lista de figuras

Ilustración 1: Pantalla principal software xPlan	3
Ilustración 2: Días libres y días festivos configurados en el software xPlan	4
Ilustración 3: Diagrama de Gantt con tareas, hitos y fechas límite ("deadlines")	5
Ilustración 4: Tareas e hitos (detalle)	6
Ilustración 5: Recursos (detalle)	7

1. Introducción al anexo de planificación

En este anexo, se detalla la planificación temporal calendarizada, con tareas e hitos, que se ha llevado a cabo para el desarrollo de este TFM.

Además, se encuentran desglosadas las tareas con sus sutareas, así como los recursos materiales que fueron utilizados para la consecución del trabajo.

Por último, también se detalla aquí, un análisis de posibles incidencias y riesgos, con un plan de contingencia y mitigación, para posibles eventualidades que pudieran haber repercutido negativamente.

Cabe aclarar, que el agregado aquí del plan de contingencia y mitigación, es sólo a efectos documentales y de metodología utilizada, ya que no ha sido necesaria la implementación del plan de contingencia y tampoco de ninguna mitigación, habida cuenta que el desarrollo se ha llevado a cabo según la idea prevista y planeada.

1.1 Planificación temporal calendarizada del Trabajo. Tareas, hitos y recursos.

1.1.1 Planificación temporal calendarizada

La cantidad de créditos del TFM es de 15 ECTS (European Credit Transfer System: Sistema acordado de cargas horarias lectivas equivalentes en los países de la UE), los cuales se traducen en una carga de 375 horas de trabajo para la consecución del mismo.

Para la realización de este TFM han sido entonces 17 semanas con un promedio de 22 horas por semana, repartidas del siguiente modo:

- 2 horas diarias de lunes a viernes = 10 horas,
- 6 horas el sábado,
- 6 horas el domingo,

Con lo cual: $10 + 6 + 6 = 22$ horas por semana.

Primera semana del proyecto: Semana 9 del calendario 2017

Última semana del proyecto: Semana 25 del calendario 2017

Total de semanas previstas para el proyecto: 17

1.1.2 Seguimiento del proyecto: utilización del software xPlan

Para el seguimiento del proyecto se utilizó el software xPlan (similar a Microsoft Project, pero para plataformas Apple Macintosh). He aquí la pantalla principal del mismo en donde se detallan las características fundamentales del proyecto:

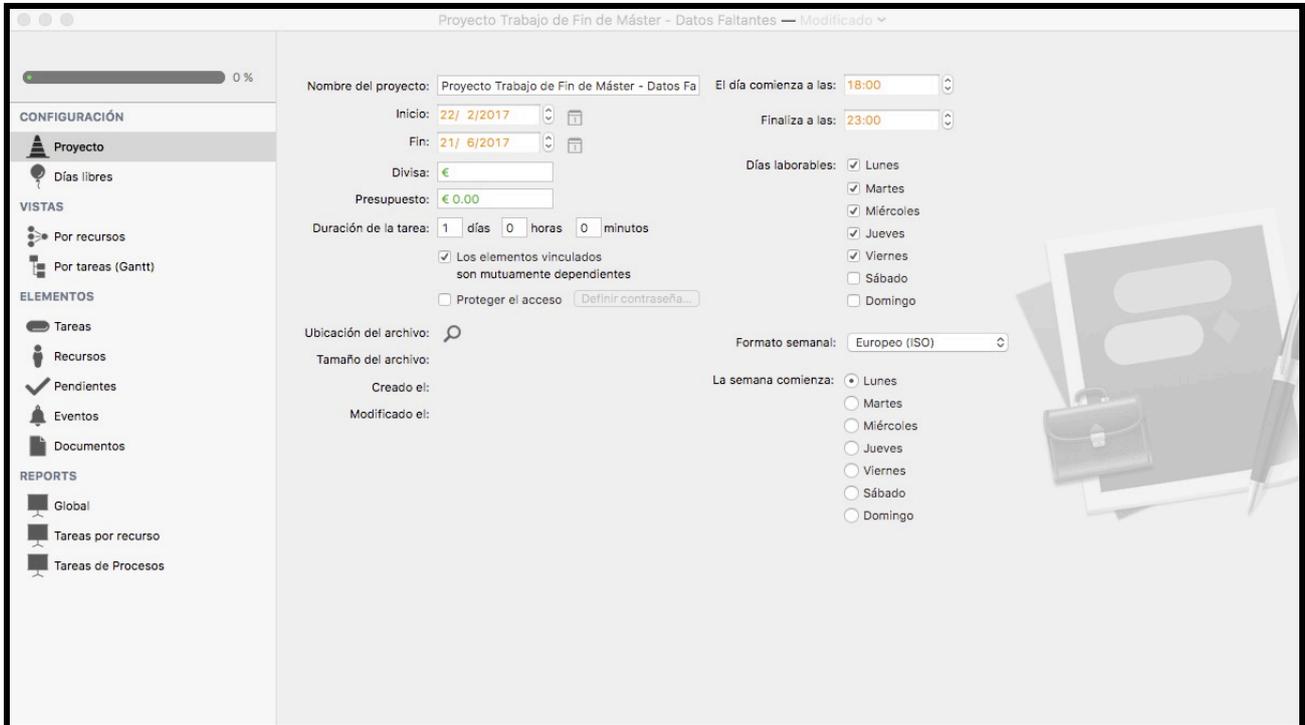


Ilustración 1: Pantalla principal software xPlan

Obsérvese como se determinó para este proyecto que los días laborables iban de lunes a viernes, desde las 18:00 h. hasta las 23:00 h. Se determinó de ese modo, ya que para la realización de este TFM, se dispuso de las horas de la tarde/noche, luego de finalizar la jornada laboral normal. Aunque en principio se planificó sólo utilizar 2 horas por día de lunes a viernes, la disponibilidad horaria fue más amplia, ya que de 18:00 h. a 23 h. se dispuso de 5 horas diarias, lo cual fue un rango de tiempo disponible que a priori fue más amplio del necesario.

Por compatibilidad con el calendario interno del programa xPlan, no se marcaron los días sábado y domingo como laborables, aunque en algún fin de semana fueron utilizados para llevar adelante este TFM, como se comentó, a razón de 6 horas el sábado y 6 horas el domingo.

1.1.3 Días libres de acuerdo al calendario de la Comunidad de Madrid:

La ciudad de residencia donde se desarrolló este TFM fue la ciudad de Madrid, por lo cual se procedió a cargar en el calendario del programa xPlan los días festivos de la Comunidad, algunos de ellos cayeron en días laborables, con lo cual, al no tener que ir a trabajar, la idea es que fueran utilizados para mitigar posibles faltas de tiempo en los días previstos a priori para realizar el desarrollo del TFM.

Desde	A	Descripción	Cada año
25/12	25/12	Navidad	<input checked="" type="checkbox"/>
1/1	1/1	Día de Año Nuevo	<input checked="" type="checkbox"/>
20/3	20/3	San José (Festivo en Madrid)	<input checked="" type="checkbox"/>
13/4	13/4	Semana Santa	<input checked="" type="checkbox"/>
14/4	14/4	Semana Santa	<input checked="" type="checkbox"/>
15/4	15/4	Semana Santa	<input checked="" type="checkbox"/>
16/4	16/4	Semana Santa	<input checked="" type="checkbox"/>
1/5	1/5	Día del Trabajador	<input checked="" type="checkbox"/>
2/5	2/5	Día de la Comunidad de Madrid	<input checked="" type="checkbox"/>
7/5	7/5	Día de la Madre	<input checked="" type="checkbox"/>
15/4	15/4	Día de San Isidro (Madrid)	<input checked="" type="checkbox"/>
25/7	25/7	Santiago Apóstol (Fest. en Mad)	<input checked="" type="checkbox"/>

Ilustración 2: Días libres y días festivos configurados en el software xPlan

1.1.4_Diagrama de Gantt

A continuación, el diagrama de Gantt del proyecto, con los enlaces de una tarea a otra tarea, dentro de las que se encontraban previstas:

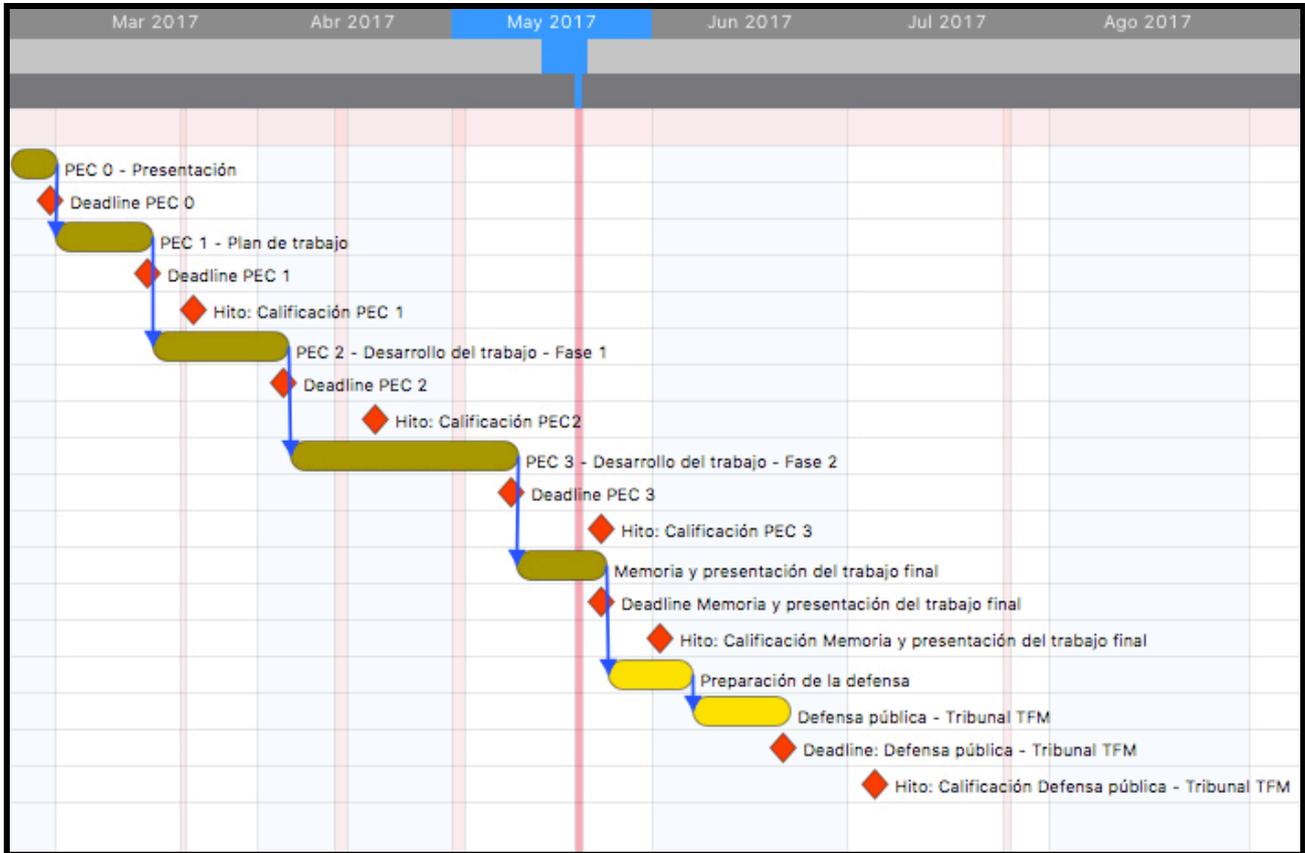


Ilustración 3: Diagrama de Gantt con tareas, hitos y fechas límite ("deadlines")

Obsérvese como cada final de actividad conecta con el inicio de la actividad siguiente, indicando el flujo de trabajo en el que ninguna tarea se solapa con otra. Los hitos (rombos de color rojo) son fechas fundamentales para la consecución del proyecto. (Véase apartado siguiente: Tareas e hitos).

Obsérvese además que la captura de pantalla visualizada corresponde al día 20 de mayo de 2017, fecha en la cual se finaliza la memoria y presentación del trabajo final de este TFM.

1.1.5 Tareas e hitos

En el siguiente esquema se pueden observar las fechas que estuvieron previstas y que fueron cumplidas, para cada una de las tareas y cada uno de los hitos del proyecto (esquema actualizado al día 20 de mayo de 2017):

	Título	Inicio	Fin	Duración	Compl...
●	PEC 0 - Presentación	22/ 2/2017 0:00	28/ 2/2017 23:00	5 días	100 %
◆	Deadline PEC 0	28/ 2/2017 0:00			
●	PEC 1 - Plan de trabajo	1/ 3/2017 0:00	15/ 3/2017 23:00	11 días	100 %
◆	Deadline PEC 1	15/ 3/2017 0:00			
◆	Hito: Calificación PEC 1	22/ 3/2017 0:00			
●	PEC 2 - Desarrollo del trabajo - Fase 1	16/ 3/2017 0:00	5/ 4/2017 23:00	14 días	100 %
◆	Deadline PEC 2	5/ 4/2017 0:00			
◆	Hito: Calificación PEC2	19/ 4/2017 0:00			
●	PEC 3 - Desarrollo del trabajo - Fase 2	6/ 4/2017 0:00	10/ 5/2017 23:00	21 días	100 %
◆	Deadline PEC 3	10/ 5/2017 0:00			
◆	Hito: Calificación PEC 3	24/ 5/2017 0:00			
●	Memoria y presentación del trabajo final	11/ 5/2017 0:00	24/ 5/2017 23:00	10 días	100 %
◆	Deadline Memoria y presentación del trabajo final	24/ 5/2017 0:00			
◆	Hito: Calificación Memoria y presentación del trabajo final	2/ 6/2017 0:00			
●	Preparación de la defensa	25/ 5/2017 0:00	6/ 6/2017 23:00	9 días	0 %
●	Defensa pública - Tribunal TFM	7/ 6/2017 0:00	21/ 6/2017 23:00	11 días	0 %
◆	Deadline: Defensa pública - Tribunal TFM	21/ 6/2017 0:00			
◆	Hito: Calificación Defensa pública - Tribunal TFM	5/ 7/2017 0:00			

Ilustración 4: Tareas e hitos (detalle)

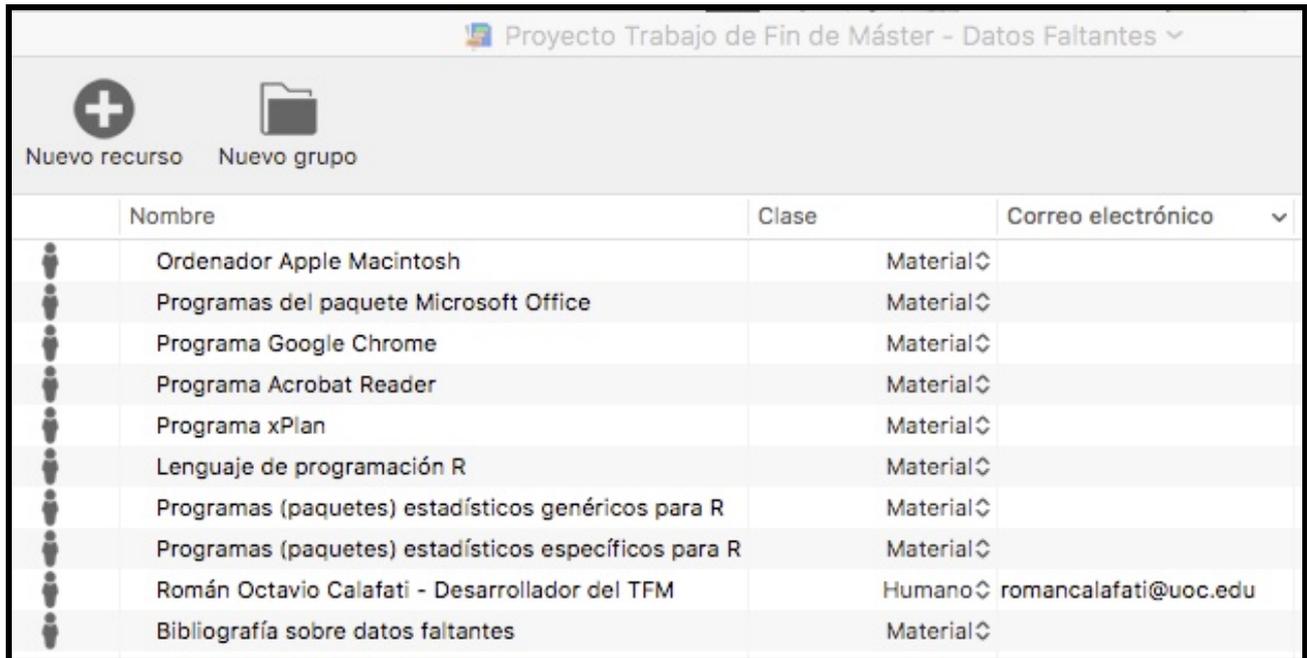
Como puede observarse en la imagen anterior (captura de pantalla del software xPlan), se consideraron como tareas las siguientes actividades:

	Inicio	Entrega	Calificación
Actividad: PEC 0 - Presentación	22/02/2017	28/02/2017	-
Actividad: PEC 1 - Plan de trabajo	01/03/2017	15/03/2017	22/03/2017
Actividad: PEC 2 - Desarrollo del trabajo - Fase 1	16/03/2017	05/04/2017	19/04/2017
Actividad: PEC 3 - Desarrollo del trabajo - Fase 2	06/04/2017	10/05/2017	24/05/2017
Actividad: Memoria y presentación del trabajo final	11/05/2017	24/05/2017	02/06/2017
Actividad: Preparación de la defensa	25/05/2017	06/06/2017	-
Actividad: Defensa pública - Tribunal TFM	07/06/2017	21/06/2017	05/07/2017

Los hitos indicados fueron las fechas límite de entrega de las tareas ("deadlines"), así como las fechas de las calificaciones de las distintas tareas. Se consideraron como hitos esas fechas, ya que en ambos casos esos hitos se ajustaban a la siguiente definición:

"Los hitos marcan los estados intermedios del proyecto y permiten avanzar en sucesivas etapas de resultados prácticos. Si se produjera un retraso en la consecución de alguno de los hitos repercutiría en otras actividades y tareas del proyecto y podría comportar un atraso en los plazos marcados. Por lo tanto, es imprescindible ser el máximo de estricto en el cumplimiento de las fechas."

1.1.6 Recursos



The screenshot shows a web interface for managing resources. At the top, it says "Proyecto Trabajo de Fin de Máster - Datos Faltantes". Below this are two buttons: "Nuevo recurso" (with a plus icon) and "Nuevo grupo" (with a folder icon). The main part of the interface is a table with the following columns: "Nombre", "Clase", and "Correo electrónico". The table contains ten rows of resources, each with a person icon on the left. The resources listed are:

	Nombre	Clase	Correo electrónico
👤	Ordenador Apple Macintosh	Material	
👤	Programas del paquete Microsoft Office	Material	
👤	Programa Google Chrome	Material	
👤	Programa Acrobat Reader	Material	
👤	Programa xPlan	Material	
👤	Lenguaje de programación R	Material	
👤	Programas (paquetes) estadísticos genéricos para R	Material	
👤	Programas (paquetes) estadísticos específicos para R	Material	
👤	Román Octavio Calafati - Desarrollador del TFM	Humano	romancalafati@uoc.edu
👤	Bibliografía sobre datos faltantes	Material	

Ilustración 5: Recursos (detalle)

Los recursos que se utilizaron para este TFM fueron:

- un ordenador Apple Macintosh con sistema operativo macOS Sierra.
- programas de software varios, tales como: Microsoft Office (Word, Excel, PowerPoint), navegador de internet Google Chrome, lector de ficheros pdf Adobe Acrobat Reader, programa de gestión de proyectos xPlan, lenguaje de programación R con paquetes estadísticos genéricos y paquetes estadísticos específicos para el tratamiento de los datos faltantes (véase el anexo 1 de este TFM, en el fichero llamado "TFM_Memoria_Final_Alumno_Roman_Octavio_Calafati").
- un recurso humano (estudiante que desarrolló este TFM).
- bibliografía específica sobre datos faltantes (véase el apartado 8 que detalla la bibliografía, en el fichero llamado "TFM_Memoria_Final_Alumno_Roman_Octavio_Calafati").

2. Incidencias, riesgos y plan de contingencia

Como se ha comentado previamente, el plan de contingencia y mitigación aquí detallado, es sólo a efectos documentales y de metodología utilizada, ya que no ha sido necesaria la implementación del mismo, habida cuenta que el desarrollo se ha llevado a cabo según la idea prevista y planificada.

2.1. Riesgo 1: mal funcionamiento del ordenador

Existe la posibilidad que durante el desarrollo del proyecto, el ordenador en el cual se trabaja, manifieste errores (de hardware o de software). Estos errores pueden originarse en incompatibilidades entre programas antiguos y programas más actuales. De todos modos, esta eventualidad tiene una probabilidad baja, ya que el ordenador en el que se trabaja tiene más de un año de uso, tiempo durante el cual no ha dado inconvenientes para el desarrollo de los trabajos que se realizaron en asignaturas anteriores.

2.1.1. Plan de contingencia y mitigación

Sin embargo, aunque el riesgo es muy bajo, dada la eventualidad de un mal funcionamiento del hardware o software del ordenador en el que se trabaja, se prevee la migración de los datos a otro ordenador que se encuentra actualmente disponible. Sólo algunos detalles menores deben ser tenidos en cuenta, tales como la conversión de algunos ficheros .txt en formato Macintosh (ya que el ordenador de trabajo opera con sistema operativo MacOS) a formato .txt de Windows (ya que el ordenador de contingencia al cual se llevarían los datos y los scripts en R, opera con sistema operativo Windows 10).

2.2 Riesgo 2: pérdida de datos (virus, malware)

Dada la eventualidad de un contagio por virus o malware en los ficheros de datos, la idea es realizar copias de respaldo cada dos o tres días de las modificaciones y adelantos que se vayan llevando a cabo en los ficheros del proyecto. Las copias de respaldo se realizan a tres destinos distintos: a) un disco duro externo portable, b) un pendrive c) una cuenta de Dropbox.

Con lo cual se cuentan con 4 copias iguales cada dos o tres días, en 4 sitios distintos. Nota: se consideran 4 copias ya que además de las 3 copias de respaldo mencionadas, obviamente existe siempre una copia de toda la información generada en el disco duro del ordenador de trabajo.

2.2.1 Plan de contingencia y mitigación

No se ha considerado un plan de contingencia para la eventual pérdida de datos por virus o malware, ya que la misma forma de trabajo con la que se actúa (4 copias en distintos sitios), elimina virtualmente la posibilidad que una pérdida de datos se de en los 4 sitios al mismo tiempo.

Para ello, deberían perderse los datos del disco duro del ordenador de trabajo, deberían perderse los datos del disco duro externo de respaldo, deberían perderse los datos del pendrive externo, y por último, deberían perderse los datos en el servidor remoto (en la nube) de Dropbox.

Si consideramos un 1% de posibilidad de pérdida de datos en cada uno de los sitios, y consideramos los 4 sitios de resguardo, se obtiene 0,01 elevado a la 4ta. potencia, ya que las probabilidades para que se produzcan los sucesos independientes de pérdidas de datos al mismo tiempo se reduce al valor 0,00000001 ($0,01^4$). Además de esta

probabilidad bajísima, hay que tener en cuenta que en sistemas Macintosh no existen prácticamente virus, así que la probabilidad de 0,00000001, es más baja en realidad.

2.3 Riesgo 3: Solapamiento con entregas en la asignatura Análisis Multivariante

La asignatura Análisis Multivariante, realizada en paralelo con el TFM, es una asignatura técnica en el área estadística que eventualmente puede generar una carga de trabajo que puede solaparse con el TFM. Se preveen 2 PECs. para esta asignatura, la primera el día miércoles 6 de abril y la segunda el día domingo 25 de junio. Como puede observarse en calendario, en ninguna de las dos fechas se produce alguna fecha límite de entrega de actividades relacionadas con el TFM, i.e., no existe solapamiento de entregables entre las dos asignaturas. Con lo cual la única posibilidad perjudicial es que la carga de trabajo en la asignatura Análisis Multivariante sea mayor de la esperada, y ocupe más horas de las previstas.

2.3.1 Plan de contingencia y mitigación

Como plan de contingencia se cuenta con las fechas de fiestas de la Comunidad de Madrid, así como también se cuenta con los días de Semana Santa. Si fuera necesario, estas fechas pueden utilizarse íntegras para trabajar en el TFM.

2.4 Riesgo 4: Carga de tiempo extra en el trabajo

En mi situación laboral actual, no hay un riesgo muy alto de tener que realizar horas extras en el trabajo.

2.4.1 Plan de contingencia y mitigación

De todos modos, dada la eventualidad, se utilizarían las fechas comentadas en el punto anterior, y también podrían solicitarse y negociarse días de vacaciones en el trabajo, los cuales serían utilizados íntegramente para el desarrollo del TFM.

2.5 Riesgo 5: Viaje imprevisto

Existe una posibilidad baja de tener que realizar un viaje por problemas de salud de algún familiar.

2.5.1 Plan de contingencia y mitigación

Se solicitarían y negociarían días de vacaciones en el trabajo, para no perjudicar al tiempo planificado para el desarrollo del TFM.

2.6 Riesgo 6: Alcance del proyecto

En cualquier proyecto existe la posibilidad que el alcance del mismo "se vaya de las manos", i.e., perder el control del mismo por el deseo de incorporar más detalles y características.

2.6.1 Plan de contingencia y mitigación

Para mitigar esta posibilidad se cuenta con el asesoramiento y experiencia de la tutora del TFM. Su consejo experto permite que el alcance del TFM no se descontrole, y se encuadre dentro de unos límites factibles en cuanto a los objetivos de alcance.

2.7 Riesgo 7: Problemas de salud personales inesperados

Puede darse la eventualidad de sufrir algún problema personal de salud inesperado durante los meses de desarrollo del TFM.

2.7.1 Plan de contingencia y mitigación

No existe un plan de contingencia y mitigación definido con exactitud para este eventualidad, ya que esto depende de las características del problema de salud, y deberá analizarse en el momento si se produjera. En el peor de los casos, se evaluaría la posibilidad de detener el desarrollo del TFM, para realizarlo en un futuro en lo posible cercano (siguiente semestre).